Quality Indicators in colonoscopy

Prof Matt Rutter
University Hospital of North Tees
UK
Possible conflicts of interest

• Nil
Overview

• Compelling evidence of the importance of quality in screening

• 10 minutes
  – Relevance of Quality Indicators
  – How to set colonoscopy Quality Indicators
  – Examples - issues with Quality Indicators & how to resolve them

  – Quality indicators also known as:
    • Quality measures
    • Performance measures
    • Key performance indicators (KPIs)
What is a Key Performance Indicator?

A measurement used to assess performance of a service/aspect of a service

- Part of the quality control cycle (Set → Measure → Act → Repeat)
- Early warning system – identifies potential problems

- KPIs help keep a balanced eye on all aspects of a service
  - Determine areas (domains) to be assessed

- KPIs should be:
  - Relevant to that domain
  - Objective & reproducible - Define KPI carefully
  - Practical & timely to measure

- Only measure things worth measuring
KPI
KPI

Minimum quality standard (QS)
KPI

- Target quality standard (QS)
- Minimum quality standard (QS)
• A quality standard may vary according to procedure
What to measure
## Colonoscopy domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Potential KPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Complete examination</td>
<td>CIR, TI intubation, rectal retroflexion, bowel prep quality</td>
</tr>
<tr>
<td>2 Accurate identification of pathology</td>
<td>ADR, MNA, PDR, MNP, PDR proximal colon, P10DR, Polypectomy rate, CDR, serrated detection rate, PCCRC, WT, bowel prep quality, lesion characterisation, biopsies in diarrhoea, chromo in IBD surveillance</td>
</tr>
<tr>
<td>3 Optimal management of pathology</td>
<td>PCCRC, Polyp Retrieval Rate, Tattooing cancers / large polyps, piecemeal resection of cancer rates, benign surgery rates, timeliness</td>
</tr>
<tr>
<td>4 Safety</td>
<td>Adverse events - Post-Polypectomy Bleeding, Perforation (Overall, Diagnostic, Therapeutic, Stricture dilatation, Stenting), Unplanned/prolonged hospital admissions, Unplanned procedure (within 8d), Deaths (within 30d), Use of reversal agents, Sedation level (Under 70; 70+)</td>
</tr>
<tr>
<td>5 Patient satisfaction</td>
<td>Comfort score, overall satisfaction with service</td>
</tr>
<tr>
<td>6 Timeliness</td>
<td>Time from referral to test, time from identification of pathology to therapy</td>
</tr>
<tr>
<td>Domain</td>
<td>Potential KPI</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1 Complete examination</td>
<td>CIR, TI intubation, rectal retroflexion, bowel prep quality</td>
</tr>
<tr>
<td>2 Accurate identification of pathology</td>
<td>ADR, MNA, PDR, MNP, PDR proximal colon, P10DR, Polypectomy rate, CDR, serrated detection rate, PCCRC, WT, bowel prep quality, lesion characterisation, biopsies in diarrhoea, chromo in IBD surveillance</td>
</tr>
<tr>
<td>3 Optimal management of pathology</td>
<td>PCCRC, Polyp Retrieval Rate, Tattooing cancers / large polyps, piecemeal resection of cancer rates, benign surgery rates, timeliness</td>
</tr>
<tr>
<td>4 Safety</td>
<td>Adverse events - Post-Polypectomy Bleeding, Perforation (Overall, Diagnostic, Therapeutic, Stricture dilatation, Stenting), Unplanned/prolonged hospital admissions, Unplanned procedure (within 8d), Deaths (within 30d), Use of reversal agents, Sedation level (Under 70; 70+)</td>
</tr>
<tr>
<td>5 Patient satisfaction</td>
<td>Comfort score, overall satisfaction with service</td>
</tr>
<tr>
<td>6 Timeliness</td>
<td>Time from referral to test, time from identification of pathology to therapy</td>
</tr>
</tbody>
</table>
Example Domain: Complete Examination
Caecal Intubation Rate (CIR)

Why measure it?

• Incomplete colonoscopy → missed pathology
  – Colonoscopist ADR correlates CIR \( r = 0.203, \ p = 0.013 \)^1
  – Colonoscopists with high completion rates are less likely to have a PCCRC ^2
  – Correlation between incomplete colonoscopy & PCCRC ^3

• Incomplete colonoscopy → subjects patient to further procedure
  – Risk, discomfort, inconvenience, expense
  – Delayed diagnosis
  – Endo unit waiting times

1. (Lee, Rutter et al. 2012)  
2. (Baxter, Sutradhar et al. 2011)  
3. (Brenner, Chang-Claude et al. 2011)
Caecal Intubation Rate (CIR)

- Should be **objective**
  - Photo of caecum (appendiceal orifice & ICV) or ileum
  - Unadjusted figure (not adjusted for poor prep/strictures – less gaming)
Example Domain: Safety
An example – post-polypectomy bleeding (PPB)

- Most frequent complication of polypectomy
- Rates of 0.26 to 6.1% $^{2-4}$

- Why the 20-fold variation?

1. (Rutter, Nickerson et al. 2013)
2. (Gavin, Valori et al. 2012)
3. (Rosen, Bub et al. 1993)
4. (Nelson, McQuaid et al. 2002)
5. (Cotton, Eisen et al. 2010)
6. (Rutter and Chilton 2011)
Why the variation?

• Differing performance?

• Differing definitions?
  – What do we mean by a bleed? Peri-procedural bleeding?
  – During that admission, or later on?
  – What severity counts? How do you stratify severity?
  – Need objective terminology, definitions & severity stratification

• Differing mechanisms for identifying complications?
  – If you don’t look, your figures will look good...

• Differing complexity of therapy?
  – Some endoscopists do advanced therapy, others do not
Why the variation?

• Differing performance?

• Differing definitions?
  – What do we mean by a bleed? Peri-procedural bleeding?
  – During that admission, or later on?
  – What severity counts? How do you stratify severity?
  ➔ Need objective terminology, definitions & severity stratification

• Differing mechanisms for identifying complications?
  – If you don’t look, your figures will look good...

• Differing complexity of therapy?
  – Some endoscopists do advanced therapy, others do not
## APPENDIX 2: REPORTING OF BLEEDING

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Severity</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectal bleeding within 30 days of procedure resulting in any of the following</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Procedure aborted</td>
<td>Minor</td>
<td>• Record on BCSS as adverse event</td>
</tr>
<tr>
<td>• Unplanned post procedure medical consultation</td>
<td></td>
<td>• Report to QARC</td>
</tr>
<tr>
<td>• Unplanned hospital admission, or prolongation of hospital stay, for $\leq 3$ nights</td>
<td></td>
<td>• Record timing post procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Record site in colorectum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Record cause of bleeding, equipment used, diathermy settings, additional factors, etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Record haemoglobin drop</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Record number of units transfused</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Record interventional procedure(s) and surgery</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Record length of stay</td>
</tr>
<tr>
<td>• Haemoglobin drop of $\geq 2$ g</td>
<td>Intermediate</td>
<td>As above, plus</td>
</tr>
<tr>
<td>• Transfusion</td>
<td></td>
<td>• Root cause analysis</td>
</tr>
<tr>
<td>• Unplanned admission or prolongation for 4–10 nights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• ITU admission for 1 night</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Intervventional procedure (endoscopic or radiological)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Surgery</td>
<td>Major</td>
<td></td>
</tr>
<tr>
<td>• Unplanned admission or prolongation for $&gt; 10$ nights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• ITU admission $&gt; 1$ night</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Death</td>
<td>Fatal</td>
<td>As above, plus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Record cause and time of death</td>
</tr>
</tbody>
</table>

(Rutter and Chilton 2011)
Why the variation?

• Differing performance?

• Differing definitions?
  – What do we mean by a bleed? Peri-procedural bleeding?
  – During that admission, or later on?
  – What severity counts? How do you stratify severity?
  – Need objective terminology, definitions & severity stratification

• Differing mechanisms for identifying complications?
  – If you don’t look, your figures will look good...

• Differing complexity of therapy?
  – Some endoscopists do advanced therapy, others do not
Why the variation?

• Differing performance?

• Differing definitions?
  – What do we mean by a bleed? Peri-procedural bleeding?
  – During that admission, or later on?
  – What severity counts? How do you stratify severity?
  – Need objective terminology, definitions & severity stratification

• Differing mechanisms for identifying complications?
  – If you don’t look, your figures will look good...

• Differing complexity of therapy?
  – Some endoscopists do advanced therapy, others do not
Post-polypectomy bleeding: \textit{PASAER-PPB}

- How to compare \textit{level 1} with \textit{level 4} endoscopist?

- Procedure-Adjusted Standardised Adverse Event Ratio (1)
  - Ratio of observed to expected adverse events
  - Standardised (weighted) according to procedure complexity
  - Automatically calculated over last 200 procedures (rolling)

- \textbf{Red}: strong evidence >2x expected rate (lower 95\% CL > 2)
- \textbf{Amber}: weaker evidence (lower 80\% CL > 2)
- \textbf{Green} otherwise

1. Blanks RG et al. \textit{Endoscopy} 2015
How to monitor performance for rare events

• Need to ensure figures reflect performance, rather than confounders

• Often better suited to qualitative review – look at each case
The practicalities
**Principles**

- Standardised & objective KPIs
- Ideally **centralised** – benchmarking very powerful

- Ethos = open & supportive
  - Share with key people – endoscopist & QA team
  - Clear processes for what to do when figures are low
Does QA work?
<table>
<thead>
<tr>
<th>Year</th>
<th>n</th>
<th>CIR</th>
<th>Comf</th>
<th>IV sed</th>
<th>Perf</th>
<th>PPB</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>5216</td>
<td>92.8</td>
<td>86.0</td>
<td>85.4</td>
<td>0.037</td>
<td>1.610</td>
</tr>
<tr>
<td>2008</td>
<td>15949</td>
<td>95.6</td>
<td>87.5</td>
<td>86.5</td>
<td>0.102</td>
<td>1.504</td>
</tr>
<tr>
<td>2009</td>
<td>25648</td>
<td>96.2</td>
<td>89.3</td>
<td>85.1</td>
<td>0.085</td>
<td>1.371</td>
</tr>
<tr>
<td>2010</td>
<td>37760</td>
<td>96.6</td>
<td>89.6</td>
<td>82.8</td>
<td>0.073</td>
<td>1.126</td>
</tr>
<tr>
<td>2011</td>
<td>45471</td>
<td>96.7</td>
<td>91.5</td>
<td>80.2</td>
<td>0.046</td>
<td>0.937</td>
</tr>
<tr>
<td>2012</td>
<td>46215</td>
<td>97.1</td>
<td>92.5</td>
<td>77.5</td>
<td>0.049</td>
<td>0.911</td>
</tr>
<tr>
<td>2013</td>
<td>37509</td>
<td>97.2</td>
<td>93.4</td>
<td>74.2</td>
<td>0.046</td>
<td>0.504</td>
</tr>
</tbody>
</table>
Summary

• Quality matters
• KPIs an important aspect of measuring quality
• Select the right domains and KPIs

• Carefully define KPIs & methodology
  – Standardised, objective

• Requires organisation
• Process should be supportive
**Withdrawal Time (WT)**

Time from colonoscope reaching caecum to removal of instrument from patient

- Endoscopists with WT > 6 min detected 2.5x more adenomas & advanced neoplasia\(^1\)

- Effect beyond 6 minutes less pronounced & mainly for non-advanced adenomas
  - Longer mean WT → increased ADR but not AA. Effect minimal beyond 10 min\(^2\)

- Adopting a minimum WT alone may not substantially improve ADR
  - Most intervention studies mandating minimum WT have been unsuccessful\(^3\)

<table>
<thead>
<tr>
<th>Table 4. Rates of Detection of Lesions According to Mean Withdrawal Time for Procedures in Which No Polyps Were Removed.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Subjects with adenomas (%)</td>
</tr>
<tr>
<td>Adenomas per subject screened (no.)</td>
</tr>
<tr>
<td>Subjects with advanced neoplasia (%)</td>
</tr>
<tr>
<td>Advanced neoplastic lesions per subject screened (no.)</td>
</tr>
<tr>
<td>Cancers per subject screened (no.)</td>
</tr>
<tr>
<td>Subjects with hyperplasia (%)</td>
</tr>
</tbody>
</table>

1. (Barclay, Vicari et al. 2006)  
2. (Lee, Blanks et al. 2013)  
3. (Corley, Jensen et al. 2011)
Patient Satisfaction
Comfort

Why measure it?
• Pain = bad technique = increased risk of complications
• Pain $\rightarrow$ patient won’t return – nor their friends & relatives $\rightarrow$ worse health outcome \(^1\)

Options
• Patient-recorded (best, but not with amnesic effect of sedation)
• Endoscopist-recorded (subject to bias)
• Nurse-recorded (less bias)
  – only 1 validated colonoscopy score \(^2\)
  – Gloucester score: unvalidated but widely used & easy

---

1. (McEntire, Sahota et al. 2013)
2. (Rostom, Ross et al. 2013)
Nurse-assessed patient comfort (Gloucester Score)

1. Comfortable
   – Talking / comfortable throughout

2. Minimal
   – 1 or 2 episodes of mild discomfort without distress

3. Mild
   – More than 2 episodes of mild discomfort without distress

4. Moderate
   – Significant discomfort experienced several times with some distress

5. Severe
   – Frequent discomfort with significant distress

Moderate/severe score correlates highly with patient-reported experience “worse than expected” (28% vs 2%; R=0.92; p<0.0001)

Ekkelenkamp, WJG 2013
Quantity, not Quality?
Number of colonoscopies per year

Why?

- Endoscopic proficiency increases with no. procedures performed \(^1\)
- Complications more common with low volume endoscopists (<300 or <200/y) \(^2,3\)

Achieving an adequate volume is essential to maintaining skills & effectively monitoring performance. Each endoscopist should perform >300PA.

A higher volume is desirable”

*(European CRC Screening Guidelines, 2010)*

- Large numbers required for accurate estimates of performance
- Endoscopists shouldn’t hide behind “confidence intervals”
  - if numbers are low & other KPIs are suboptimal, action is required

1. (Enns 2007)
2. (Rabeneck, Paszat et al. 2008)
3. (Singh, Penfold et al. 2009)
10 years ago

CIR 77%
more objective – 57%

Bowles, Gut 2004
English NHS colonoscopy

10 years ago

CIR 77%
more objective – 57%

Now:
screening service

CIR 97%
unadjusted - objective

BCSP stats, 2013
Screening Quality Assurance

• 58 screening centres
  • Must be JAG-accredited

• Dataset web-entered during procedure by screening nurse onto single national database

• Open QA data
  • Updated daily

• National benchmark
• Regional QA leads oversee
Notes

- ESGE QIC work
- Why use KPIs (quality ethos etc – ESGE paper)
- Make sure I use QM or KPI throughout
- Give some examples of QS – e.g. 95% for CIR
- What are the issues/down-sides – inc gaming, time, cost, missing the point
- How to monitor (ESGE paper – centralise etc)
- What can be done if KPIs suboptimal
- Cover fact that ADRs can rarely be compared due to different populations (PCCRC rate irony?)
- The proposed title of your talk is “Quality indicators in colonoscopy” in Session 4: Screening programs and guidelines which is currently scheduled to start at 4.40 pm. The talk should be 10 minutes in length. A preliminary program is attached, please note that this is not finalized, we hope to complete and circulate it to the membership within ten days.
- The audience is international and very well informed, so speakers should directly address key issues with the group.
- Should you have any questions or comments about the proposal please feel free to get in touch directly with Professor Kuipers, email: e.j.kuipers@erasmusmc.nl or the secretariat.
- After the talks, all speakers of the respective session are asked to take a seat at the chairman’s table on the stage. The session will end with discussion and speakers are encouraged to respond to questions from the audience.
Accurate Identification of Pathology
<table>
<thead>
<tr>
<th>Performance indicator</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenoma Detection Rate (ADR)</td>
<td>Highly clinically relevant</td>
<td>Difficult to measure Susceptible to “one and done” Varies with population/age/gender</td>
<td>Lots</td>
</tr>
<tr>
<td>Mean no. Adenomas per Procedure (MNA)</td>
<td>Not susceptible to “one and done”</td>
<td>Skewed by patients with 10+ adenomas</td>
<td>Greater spread between endoscopists (Barclay NEJM 2006)</td>
</tr>
<tr>
<td>Polyp Detection Rate (PDR)</td>
<td>Easy to measure – avoids link with pathology</td>
<td>Susceptible to gaming</td>
<td>High correlation with ADR (Francis GIE 2011)</td>
</tr>
<tr>
<td>Mean no. Polyps per Procedure (MNP)</td>
<td>As for MNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polypectomy rate (PR)</td>
<td>Less susceptible to gaming</td>
<td></td>
<td>PRs of 40/30% correlate with ADRs of 25/15% for M/F (Williams GIE 2012)</td>
</tr>
<tr>
<td>PDR proximal</td>
<td>Less susceptible to gaming</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDR for 10+mm polyps (P10DR)</td>
<td>Higher correlation to adenoma detection</td>
<td>May encourage size gaming</td>
<td>84% specific for advanced neoplasia (Lieberman JAMA 2008)</td>
</tr>
<tr>
<td>Serrated Polyp Detection Rate (SPDR)</td>
<td>Clinically relevant</td>
<td>Need link to pathology Too early to know appropriate rates, Dependant on pathologist’s threshold/knowledge</td>
<td></td>
</tr>
<tr>
<td>Cancer Detection Rate (CDR)</td>
<td>Clinically relevant</td>
<td>Relates more to patient population than endoscopist</td>
<td></td>
</tr>
<tr>
<td>Post-Colonoscopy CRC rate (PCCRC)</td>
<td>Highly clinically relevant Relates to endoscopist not patient population</td>
<td>Difficult to capture</td>
<td>Time delay</td>
</tr>
<tr>
<td>Performance indicator</td>
<td>Advantages</td>
<td>Disadvantages</td>
<td>Evidence</td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Adenoma Detection Rate (ADR)</td>
<td>Highly clinically relevant</td>
<td>Difficult to measure</td>
<td>Lots</td>
</tr>
<tr>
<td>Mean no. Adenomas per Procedure (MNA)</td>
<td>Not susceptible to “one and done”</td>
<td>Skewed by patients with 10+ adenomas</td>
<td>Greater spread between endoscopists (Barclay NEJM 2006)</td>
</tr>
<tr>
<td>Polyp Detection Rate (PDR)</td>
<td>Easy to measure – avoids link with pathology</td>
<td>Susceptible to gaming</td>
<td>High correlation with ADR (Francis GIE 2011)</td>
</tr>
<tr>
<td>Mean no. Polyps per Procedure (MNP)</td>
<td>As for MNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polypectomy rate (PR)</td>
<td>Less susceptible to gaming</td>
<td></td>
<td>PRs of 40/30% correlate with ADRs of 25/15% for M/F (Williams GIE 2012)</td>
</tr>
<tr>
<td>PDR proximal</td>
<td>Less susceptible to gaming</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDR for 10+mm polyps (P10DR)</td>
<td>Higher correlation to adenoma detection</td>
<td>May encourage size gaming</td>
<td>84% specific for advanced neoplasia (Lieberman JAMA 2008)</td>
</tr>
<tr>
<td>Serrated Polyp Detection Rate (SPDR)</td>
<td>Clinically relevant</td>
<td>Need link to pathology, Too early to know appropriate rates, Dependent on pathologist’s threshold/knowledge</td>
<td></td>
</tr>
<tr>
<td>Cancer Detection Rate (CDR)</td>
<td>Clinically relevant</td>
<td>Relates more to patient population than endoscopist</td>
<td></td>
</tr>
<tr>
<td>Post-Colonoscopy CRC rate (PCCRC)</td>
<td>Highly clinically relevant</td>
<td>Difficult to capture, Relates to endoscopist not patient population</td>
<td></td>
</tr>
<tr>
<td>Performance indicator</td>
<td>Advantages</td>
<td>Disadvantages</td>
<td>Evidence</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Adenoma Detection Rate (ADR)</td>
<td>Highly clinically relevant</td>
<td>Difficult to measure</td>
<td>Lots</td>
</tr>
<tr>
<td>Mean no. Adenomas per Procedure (MNA)</td>
<td>Not susceptible to “one and done”</td>
<td>Skewed by patients with 10+ adenomas</td>
<td>Greater spread between endoscopists (Barclay NEJM 2006)</td>
</tr>
<tr>
<td>Polyp Detection Rate (PDR)</td>
<td>Easy to measure – avoids link with pathology</td>
<td>Susceptible to gaming</td>
<td>High correlation with ADR (Francis GIE 2011)</td>
</tr>
<tr>
<td>Mean no. Polyps per Procedure (MNP)</td>
<td>As for MNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polypectomy rate (PR)</td>
<td>Less susceptible to gaming</td>
<td></td>
<td>PRs of 40/30% correlate with ADRs of 25/15% for M/F (Williams GIE 2012)</td>
</tr>
<tr>
<td>PDR proximal</td>
<td>Less susceptible to gaming</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDR for 10+mm polyps (P10DR)</td>
<td>Higher correlation to adenoma detection</td>
<td>May encourage size gaming</td>
<td>84% specific for advanced neoplasia (Lieberman JAMA 2008)</td>
</tr>
<tr>
<td>Serrated Polyp Detection Rate (SPDR)</td>
<td>Clinically relevant</td>
<td>Need link to pathology</td>
<td></td>
</tr>
<tr>
<td>Cancer Detection Rate (CDR)</td>
<td>Clinically relevant</td>
<td>Relates more to patient population than endoscist</td>
<td></td>
</tr>
<tr>
<td>Post-Colonoscopy CRC rate (PCCRC)</td>
<td>Highly clinically relevant</td>
<td>Difficult to capture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relates to endoscopist not patient population</td>
<td>Time delay</td>
<td></td>
</tr>
</tbody>
</table>
How do you know the quality of a service?
You don’t, unless you *measure it*